

EV316936512

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

**VIDEO-BASED GAIT RECOGNITION**

Inventor:

**Ross G. Cutler**

ATTORNEY'S DOCKET NO. MS1-1504US

1 This application claims the benefit of Provisional Application No. 60/470,670, filed  
2 on May 15, 2003, which is incorporated herein by reference in its entirety.

### 3 4 **TECHNICAL FIELD**

5 This invention relates to the analysis of the motion of a human subject, and, in a  
6 more particular implementation, to the analysis of the gait of the human subject based on  
7 video data, and the use of the analyzed gait to identify the human subject.

### 8 9 **BACKGROUND**

10 Gait refers to a person's walking style (that is, the "way" the person walks).  
11 There is strong evidence from psychophysical experiments and gait analysis research (a  
12 multi-disciplinary field that spans kinesiology, physiotherapy, orthopedic surgery,  
13 ergonomics, etc.) that gait dynamics contain a signature that is characteristic of, and  
14 possibly unique to, each individual. More specifically, from a biomechanics standpoint,  
15 human gait consists of synchronized, integrated movements of hundreds of muscles and  
16 joints of the body. These movements follow the same basic bipedal pattern for all  
17 humans, and yet vary from one individual to another in certain details (such as their  
18 relative timing and magnitudes) as a function of their entire musculo-skeletal structure,  
19 e.g., body mass, limb lengths, bone structure, etc. Because this structure is difficult to  
20 replicate, gait is believed to be unique to each individual and can be characterized by a  
21 few hundred kinematic parameters, namely the angular velocities and accelerations at  
22 certain joints and body landmarks.

23 Various techniques have been proposed for automatically analyzing a person's  
24 gait for use in different applications. One such application is gait recognition. In the  
25 computer vision community, gait recognition refers to the task of automatically extracting

1 visual cues that characterize the motion of a walking person from video images of the  
2 walking person, and using these cues to potentially identify the person. Gait is an  
3 attractive biometric for automated recognition, particularly for passive surveillance  
4 applications, due to the ability to determine this biometric "at a distance," that is, without  
5 the need to interact with the subject, or even obtain the cooperation of the subject. Gait is  
6 also a biometric that may be difficult to conceal.

7 Existing automated approaches to analyzing human gait can be categorized as  
8 either model-based or holistic. Model-based approaches use a model of either the  
9 person's shape (e.g., structure) or motion in order to recover features of gait mechanics,  
10 such as stride dimensions and kinematics of joint angles. In holistic techniques, gait is  
11 characterized by the statistics of the spatiotemporal patterns generated by the silhouette of  
12 the walking person in the image. That is, holistic techniques compute a set of features  
13 (the gait signature) from these patterns, and then use these features to classify gait. Some  
14 studies have required the subject to wear special instruments (such as LED markers), or  
15 walk on special surfaces. Other studies have attempted to characterize a person's gait  
16 without any such artificial cues and assumptions.

17 While the above-identified techniques have shown promise in identifying  
18 individuals based on their gait, there remains room for significant improvement in this  
19 field. More specifically, because of the complexity of human locomotion, it is not an  
20 intuitive matter what collection of parameters associated with a subject's ambulatory  
21 motion can best be used to characterize the subject's unique gait. It is likewise a non-  
22 routine task to provide a reliable technique for extracting identified parameters from  
23 video data. Various real-world conditions may further compound these challenges, such  
24 as the possibility that various factors may impact the subject's gait behavior in ways that  
25 may be difficult to predict (including factors of fatigue of the subject, mood of the

1 subject, sickness of the subject, footwear used by the subject, and so on), the possibility  
2 that environmental conditions may degrade the quality of the captured images, the  
3 possibility that the subject's apparel may obscure the gait behavior, the possibility that  
4 the video itself may be of relatively low resolution, and so on. These factors may result  
5 in gait analysis results that have a considerable degree of error.

6 As such, there is an exemplary need in the art to provide motion-based biometric  
7 analysis having improved accuracy, reliability, utility, and/or efficiency compared to  
8 known techniques.

## 10 **SUMMARY**

11 Described herein is a method for identifying an ambulatory subject based on  
12 video analysis of the subject's gait. The method includes: acquiring a sequence of  
13 images that collectively captures the gait of the subject; determining a collection of  
14 features from the sequence of images representative of the gait; and identifying the  
15 subject based on the determined features. In one exemplary implementation, the subject  
16 is identified based on the features of cadence, stride length, and height parameters (where  
17 the height parameters model a dynamically varying ambulatory subject height).

18 A related apparatus is also described herein.

## 20 **BRIEF DESCRIPTION OF THE DRAWINGS**

21 Fig. 1 shows an overview of an exemplary system for analyzing the gait of a  
22 human subject and identifying the human subject in response thereto.

23 Fig. 2 shows a video image of a walking subject, where the subject is in a so-  
24 called fronto-parallel orientation with respect to a video camera.

1 Fig. 3 shows a video image of a walking subject, where the subject is in a non-  
2 fronto-parallel orientation with respect to the video camera.

3 Fig. 4 shows an exemplary technique for identifying a subject based on the gait-  
4 related features of stride and cadence.

5 Fig 5 shows three versions of a video image showing an ambulatory subject for  
6 use in explaining the extraction of foreground information from background information  
7 in the video image.

8 Fig. 6 shows a sequence of binary profiles of a walking subject, and also shows  
9 the use of bounding boxes used to enclose the binary profiles.

10 Fig. 7 shows graphs used to derive the period of the subject's gait based on the  
11 width of the bounding boxes (shown in Fig. 6) as a function of time. The graphs pertain  
12 to a walking sequence that reflects a fronto-parallel orientation of the subject with respect  
13 to the camera.

14 Fig. 8 shows other graphs used to derive the period of the subject's gait based on  
15 the width of the bounding boxes (shown in Fig. 6) as a function of time. The graphs  
16 pertain to a walking sequence that reflects a non-fronto-parallel orientation of the subject  
17 with respect to the camera.

18 Fig. 9 shows a mapping of measured stride length ( $L$ ) and cadence( $C$ ), particularly  
19 illustrating the linear relationship between stride length and cadence.

20 Fig. 10 shows another exemplary technique for identifying the subject based on  
21 gait-related features of stride, cadence, and apparent height.

22 Fig. 11 shows the orientation of the camera with respect to the subject for three  
23 different identified cases pertinent to the computation of the apparent height of the  
24 subject.  
25

1 Figs. 12-14 show the geometrical parameters relevant to the computation of  
2 apparent height for three different cases that are illustrated in Fig. 14.

3 Fig. 15 shows a simplified representation of the subject's profile, illustrating the  
4 measurement of image height in a non-fronto-parallel orientation.

5 Fig. 16 shows the application of the height measurement technique shown in Fig.  
6 15 to a series of binary profiles reflecting a non-fronto-parallel orientation.

7 Fig. 17 shows a magnified view of a binary profile in a non-fronto-parallel  
8 orientation, showing the application of the measurement technique shown in Fig. 15 to  
9 measure image height.

10 Fig. 18 shows the application of the height measurement technique shown in Fig.  
11 15 to a series of binary profiles reflecting a fronto-parallel orientation.

12 Fig. 19 presents a graph that shows the relation between measured height and a  
13 model used to approximate the measured height.

14 Fig. 20 shows a technique for identifying the subject using a holistic gait-  
15 recognition approach.

16 Fig. 21 shows a computed self-similarity plot for the case of the fronto-parallel  
17 orientation.

18 Fig. 22 shows another computed self-similarity plot for the case of the non-fronto-  
19 parallel orientation.

20 Fig. 23 shows the application of units of self-similarity to a self-similarity plot.

21 Fig. 24 presents a technique for identifying the subject using an integration of the  
22 parametric-based approach and the holistic approach.

23 Fig. 25 presents another technique for identifying the subject using an integration  
24 of the parametric-based approach and the holistic approach.  
25

1 The same numbers are used throughout the disclosure and figures to reference like  
2 components and features. Series 100 numbers refer to features originally found in Fig. 1,  
3 series 200 numbers refer to features originally found in Fig. 2, series 300 numbers refer  
4 to features originally found in Fig. 3, and so on.

## 5 6 **DETAILED DESCRIPTION**

### 7 **A. System Overview**

8 A technique and related system are described herein for analyzing a motion-based  
9 biometric of a human subject, and for identifying the human subject based on the motion-  
10 based biometric. In the implementation described below, the motion-based biometric  
11 pertains to the gait of the human subject. The gait of the subject loosely refers to the way  
12 that the subject walks, which, in turn, may be based on a host of physiological  
13 parameters. The subject's gait may be unique to the subject, and in any event, has the  
14 potential of discriminating the subject from a group of other subjects having their own  
15 respective distinctive gaits.

16 The phrase "identifying the gait" has broad connotation as used here. In one case,  
17 the term "identifying" may refer to determining the identity of the walking subject (that  
18 is, determining who the walking subject is). In another case, the term "identifying" may  
19 refer to determining some characteristic of the walking subject based on their gait besides  
20 their identity.

21 The identification technique described here likewise has broad application to  
22 different fields. In one case, the technique can be used for video surveillance. For  
23 instance, the technique can be used to determine the identity of individuals by comparing  
24 these individuals' walking styles with prior stored knowledge regarding their respective  
25 walking styles. This determination can be used to detect the presence of authorized or

1 unauthorized persons in a defined area (such as a building, courtyard, etc.), to pick out a  
2 particular subject of interest in a group of other individuals, to track such a subject of  
3 interest throughout the day, and so on. The camera used for these applications can  
4 comprise a conventional video camera (either fixed or moving, analog or digital),  
5 telescopic type image acquisition devices (for observing subjects at a relatively great  
6 distance), satellite imaging acquisition devices, and so on.

7 In another application, the identification technique can be used to perform  
8 indexing of video frames for use in video playback or video editing. For instance, this  
9 technique can be used to find the beginning of a scene that included a specified  
10 ambulatory subject, etc. This technique could also serve as a video retrieval strategy.  
11 That is, the technique can be used for retrieving an identified video resource containing a  
12 specified ambulatory subject from an archive of video resources.

13 In another application, the identification technique can be used for various kinds  
14 of man-machine interfaces. For instance, the technique can be used to determine the  
15 behavior of a subject that is interacting with a system (whether knowingly or  
16 unwittingly). For instance, an exemplary application of this type can determine the  
17 identity of an individual approaching or moving away from a system based on their gait.  
18 This analysis can be used to open or lock a door or gate based on an assessment of the  
19 identity of the subject, provide various resources (such as activating an account  
20 associated with an approaching subject), or take any other actions depending on the  
21 context of the particular application.

22 In another application, the identification technique can also be used to provide  
23 information regarding the characteristics of the subject's gait. For instance, the technique  
24 can be used to identify various abnormalities in the subject's gait, potentially suggesting  
25 injury, sickness, or simply the formation of poor behavioral habits. The identification



1 technique can also be used to assess the ambulatory performance of the subject, such as  
2 by providing assessment of sports-related performance, arts-related performance (e.g.,  
3 dance performance), efficiency-related work performance, and so on.

4 The technique can be applied to still additional environments, as will be apparent  
5 to those skilled in the art.

6 Fig. 1 shows an overview of an exemplary system 100 for analyzing the gait of a  
7 human subject 102 (henceforth "subject") and identifying the subject 102 in response  
8 thereto. The subject's 102 movement over a surface defines a path 104. The path 104  
9 can represent a course over a flat surface, over an inclined surface, or generally over an  
10 uneven surface. In any case, it is assumed herein that the subject 102 is walking in an  
11 upright position in a natural fashion. Further, in one exemplary and non-limiting  
12 implementation, it is assumed that the subject 102 is walking with an approximately  
13 constant velocity (e.g., in both speed and direction) for about 5-10 seconds (e.g., which  
14 may correspond to about 10-20 steps).

15 A camera 106 captures video data that describes the subject's 102 movement.  
16 The gait-analysis techniques described herein can accommodate different orientations of  
17 the camera 106 with respect to the subject 102. In one application, the axis formed by the  
18 direction that the camera 106 is pointed in (i.e., the optical axis) intersects the path 104 of  
19 the subject 102 at an approximately right angle. This orientation is referred to as a  
20 fronto-parallel orientation herein. Orientations which deviate from the fronto-parallel  
21 orientation to a significant degree are referred to as non-fronto-parallel orientations. In  
22 Fig. 1, the illustrated orientation of the camera 106 with respect to the subject 102 can be  
23 regarded as somewhat non-fronto-parallel, since the direction that the camera 106 is  
24 pointed in varies somewhat from the normal of the path 104.  
25

1 More specifically, jumping ahead briefly in the sequence of figures, Fig. 2 shows  
2 one example of a near fronto-parallel orientation of the camera 106 with respect to a  
3 subject 202, and Fig. 3 presents one example of a non-fronto-parallel orientation of the  
4 camera 106 with respect to a subject 302. As apparent from the examples shown in Figs.  
5 2 and 3, the camera 106 can be mounted at some elevated position with respect to the  
6 walking surface, but can also be located on a position below the walking surface, or on  
7 the walking surface. The camera 106 is preferably calibrated with respect to the ground  
8 plane (e.g., walking surface) that the subject 102 is traversing. The camera 106 can be  
9 stationary or can move (e.g., in a prescribed scan pattern). In the later case, logic can be  
10 used to process the output of the camera 106 to account for the movement of the camera  
11 106.

12 Performance of gait analysis is typically optimal for the fronto-parallel  
13 orientation. But the techniques described herein also provide satisfactory results for non-  
14 fronto-parallel orientations. In the context of surveillance, it is desirable to make the gait  
15 analysis technique robust to different kinds of orientations of the camera 106 with respect  
16 to the subject 102, because the subject 102 is not under control of the entity conducting  
17 the surveillance, and therefore cannot be expected to tailor their walking so that it  
18 conforms to ideal image acquisition conditions.

19 Returning to Fig. 1, the camera 106 itself can comprise any kind of apparatus for  
20 capturing a sequence of images that define the movement of the subject 102. For  
21 instance, the camera 106 can comprise a conventional video camera. The camera 106 can  
22 capture video data 108 defining a series of video frames. Each of these video frames  
23 captures a snapshot of the subject's 102 motions at a particular juncture along the path  
24 104. The camera 106 can originally capture digital video data, or, in an alternative  
25 implementation, can capture analog video data and then convert this analog data to digital

1 form so that it can be suitably processed by digital processing devices. As mentioned  
2 above, the camera 106 can alternatively represent a telescopic image-acquisition device, a  
3 satellite acquisition device, or other kind of image acquisition device. In any case, in one  
4 exemplary application, the preferred frame rate of the camera 106 is greater than twice  
5 the walking frequency of the subject 102 (to be described in greater detail below). The  
6 camera 106 can provide video data having relatively low resolution, or video data having  
7 relatively high resolution (and, of course, any resolution level in between). For instance,  
8 in one illustrative case, the camera 106 provides image frames at 30 fps (frames per  
9 second) each with an image size of 360 x 240 pixels. In one exemplary implementation,  
10 the camera 106 can be configured to detect images in the visible portion of the  
11 electromagnetic spectrum. In other exemplary implementations, the camera 106 can be  
12 configured to detect images in other portions of the electromagnetic spectrum (such as  
13 the infrared portion), or using other imaging techniques.

14 A processing module 110 receives the video data 108 via coupling path 112. The  
15 coupling path 112 can represent a hardwired link or a wireless (e.g., radio) link between  
16 the camera 106 and processing module 110. That coupling path 112 can be formed as a  
17 point to point coupling path, or can be formed using any kind of network (Internet,  
18 intranet, LAN, etc.). Alternatively, the path 112 can loosely represent the manual transfer  
19 of video data via transportable media (e.g., magnetic or optical media, etc.).

20 The camera 106 can be located locally with respect to the processing module 110.  
21 Alternatively, the camera 112 can be positioned at a remote location with respect to the  
22 processing module 110. In still another case, the camera 106 can be integrated with the  
23 processing module 110; for example, a video camera can include integrated processing  
24 logic for performing analysis on the captured video, either in real time fashion, or upon  
25 playback of recorded video. The camera 106 and processing module 110 can be under

1 the control of the same agent (e.g., controlling person or business entity), or can be under  
2 the control of different agents (such as when a first entity makes a video recording of a  
3 walking subject, and a second entity accesses this video recording and perform gait  
4 analysis on this video recording).

5 In any case, the processing module 110 can perform real-time or near real-time  
6 analysis of the subject's 102 gait in a real-time mode, as well as a non-real-time analysis  
7 in a non-real-time mode. In other words, in the real-time mode, the system 100 can be  
8 configured to immediately forward captured video data 108 to gait analysis logic 114 for  
9 immediate execution of gait analysis thereon. The real-time mode of operation is  
10 represented by path 116 that feeds video data from the camera 106 directly to the gait  
11 analysis logic 114. Alternatively, in the non-real-time mode, the system 100 can be  
12 configured to forward the video data 108 to a data storage 118. Then, at some later time,  
13 the processing module 110 can be configured to retrieve the video data 108 from the data  
14 storage 118 and forward it to the gait analysis logic 114 for performing analysis thereon.  
15 The non-real-time mode of operation is represented by path 120 in Fig. 1 that shows the  
16 video data 108 being forwarded from the data storage 118 to the gait analysis logic 114.  
17 The data storage 118 can represent a memory storage device that can be integrated with  
18 the gait analysis logic 114 (e.g., located in the same physical housing as the gait analysis  
19 logic 114), or can be located separate from the gait analysis logic 114 (e.g., as in the case  
20 where the data storage 118 is coupled to the gait analysis logic 114 via point-to-point  
21 connection or via network, etc.).

22 The processing module 110 itself can comprise any kind of processing  
23 functionality, including any kind of general purpose or special purpose computer, neural  
24 network, etc. To simply the explanation, it is assumed in the context of this discussion  
25 that the processing module 110 represents a computer. In this case, the computer can

1 represent a stand-alone computer workstation, a server-type computer coupled to a  
2 network, or any other kind of computing device. Although not illustrated,, such a  
3 computer can include conventional computer hardware, such as memory (ROM, RAM,  
4 magnetic discs, optical discs, etc.), one or more processing devices, one or more  
5 communication interfaces, one or more user input devices (keyboard, mouse device, etc.),  
6 one or more user output devices (display monitor, printer, etc.), and one or more buses  
7 for coupling its components together. Alternatively, as described above, the processing  
8 module 110 can represent logic functionality that is integrated with the camera 106, and,  
9 in this case, can represent computing hardware and software specifically tailored for its  
10 intended use in the camera 106.

11       However constituted, the memory (not shown) of the processing module 110 can  
12 include application logic for perform prescribed functions. Such application logic can  
13 take the form a series of machine instructions stored in memory for performing the  
14 ascribed functions. In the context of the present discussion, the application logic can  
15 include the above-mentioned gait analysis logic 114 for analyzing the sequence of frames  
16 in the video data 108, analyzing the gait-related behavior of the subject captured by the  
17 video data 108, and performing classification or discrimination functions based on the  
18 assessed gait-related behavior.

19       More specifically, the gait-related logic 114 can include a series of logic modules  
20 that perform individual subtasks within the general analysis task. Each of these modules  
21 can represent a collection of machine instructions for performing the module's ascribed  
22 subtask, such as a subroutine for performing the module's ascribed subtask.

23       Such modules can include preprocessing logic 122. The preprocessing logic 122  
24 examines the video data 108 captured by the camera 106 and transforms the video data  
25 108 into a form suitable for analysis by the remainder of the modules in the gait analysis

1 logic 114. For instance, the preprocessing logic 122 can include logic for extracting the  
2 portions of the video frames that represent the walking subject 102 from a larger field of  
3 video information (e.g., representing the background in the image scene).

4 Feature extraction logic 124 extracts one or more parameters associated with gait  
5 from the preprocessing video data. If configured to operate in a parametric mode, the  
6 feature extraction logic 124 extracts one or more parameters based on a model (or  
7 multiple models) of the subject's 102 gait behavior. If configured to operate in a non-  
8 parametric (holistic) mode, the feature extraction logic 124 identifies features extracted  
9 from a spatiotemporal consideration of the video data itself, without necessarily making  
10 reference to an explicit theoretical model that describes gait behavior. In other words, the  
11 parametric analysis typically extracts targeted data from the video data 108 for analysis,  
12 whereas the holistic approach typically considers the video data 108 as a whole, thus  
13 deriving features based on an aggregative or statistical analysis of the gait behavior  
14 contained within the video data 108 as a whole. In still another application, the feature  
15 extraction logic 124 can combine aspects of both parametric and holistic approaches.

16 Classification logic 126 receives the features identified in the feature extraction  
17 logic 124 and determines some attribute of the subject 102 based on the extracted  
18 features. In one case, the classification logic 126 can compare the extracted features with  
19 previously encountered gait information corresponding to different individuals to  
20 determine the identity of a subject 102 under consideration. That is, in this application,  
21 the classification logic 126 can be used to match a "gait signature" of the subject under  
22 consideration with one or more gait signatures contained in a pool of such signatures  
23 stored in a classification knowledgebase. The classification logic 126 can also include  
24 functionality for training its discrimination functionality such that this functionality  
25 becomes progressively more accurate as additional video data is processed. One type of

1 analysis technique that can be used to perform the above-identified tasks is the K nearest  
2 neighbor classification technique. Further, principal component analysis can be used to  
3 cull a salient grouping of information from the output of the feature extraction logic 124,  
4 and to present this salient grouping of information to the classification logic 126 instead  
5 of the original (e.g., raw) feature data generated by the feature extraction logic 124.

6 The following sections provide additional details regarding the three modules  
7 identified above, namely the preprocessing logic 122, the feature extraction logic 124,  
8 and the classification logic 126. More specifically, Section B describes a technique for  
9 recognizing subjects using a parametric approach based on the features of stride and  
10 cadence (to be defined below). Section C describes a technique for recognizing subjects  
11 using a parametric approach based on the features of stride, cadence and apparent height  
12 (where apparent height is measured as a dynamically changing parameter). Section D  
13 describes a technique for recognizing subjects using a holistic approach. Finally, Section  
14 E describes a technique for recognizing subjects using a combination of parametric and  
15 holistic approaches.

#### 16 17 B. Technique for Recognizing Subjects Using Linearly-Related Parameters of 18 Stride and Cadence

19 Fig. 4 shows an exemplary method 400 for identifying the subject 102 based on  
20 the gait-related features of stride and cadence (to be defined below). The method 400  
21 includes three general steps corresponding to the logic modules provided in the gait  
22 analysis logic 114 of Fig. 1. Namely, in a first general step 402, video data 108 is  
23 preprocessed to transform it into a form suitable for feature extraction. In the second  
24 general step 404, the transformed video data 108 is processed to extract stride and  
25 cadence parameters from the transformed data. In a third general step 406, the extracted

1 features are used to classify the subject 102, or otherwise derive some attribute of the  
2 subject 102. Each of the above-described general steps will be described in further detail  
3 below.

4 The first general step, step 402, includes a first substep 408 that entails modeling  
5 background information in the video data 108, a second substep 410 that entails  
6 segmenting moving objects in the video data 108, and a third substep 412 that entails  
7 tracking the subject 102 within the video data 108.

8 More specifically, substeps 408 and 410 generally involve identifying background  
9 information in the scene represented by the video data 108 (such as information  
10 corresponding to the pavement, grass, and building shown in Fig. 3), identifying  
11 foreground information in the scene (such as the moving subject 302 shown in Fig. 3),  
12 and extracting the foreground information from the background information. Different  
13 known techniques are available for performing these tasks. In one exemplary  
14 implementation, a non-parametric background modeling technique is used to detect  
15 foreground information in the video data 108. This technique is robust to lighting  
16 changes, camera jitter, and to the presence of shadows. One such non-parametric  
17 technique is described in A. Elgammal, D. Harwood, and L. S. Davis, "Non-parametric  
18 Model for Background Subtraction," IEEE ICCV99 Frame Rate Workshop, IEEE 7<sup>th</sup>  
19 International Conference on Computer Vision, Kerkyra, Greece, September 1999.

20 Fig. 5 shows an exemplary series of images that illustrate the operation of  
21 substeps 408 and 410. Namely, image 502 in Fig. 5 represents an original image  
22 containing a walking subject 504, where the camera defines a fronto-parallel orientation  
23 with respect to the subject 504. Image 506 shows a foreground template produced from  
24 the original image 502, which shows the subject 504, but omits the background  
25 information. Image 508 shows a binary template (also referred to as a binary silhouette



1 herein) that shows the distinction between background and foreground information by  
2 presenting these regions in black and white, respectively (e.g., 1 and 0, respectively).  
3 Alternatively, the binary silhouette can be produced by presenting the background and  
4 foreground regions in white and black, respectively.

5       Once the subject 102 has been detected in a video frame, substep 412 comes into  
6 play by tracking the subject 102 in subsequent frames. Again, known techniques are  
7 available for performing this ascribed task. In one exemplary technique, binary  
8 silhouettes are tracked from frame to frame in the video data 108 using spatial and  
9 temporal coherence techniques. For instance, tracking can be performed in a simple case  
10 by linking overlapping binary silhouettes that appear in multiple respective frames. One  
11 exemplary technique that can be used is described in I. Haritaoglu, D. Harwood, and L. S.  
12 Davis, "W4s: A Real-Time System for Detecting and Tracking People in 21/2 d," in  
13 European Conference on Computer Vision, 1998. A walking subject 102 can be  
14 discriminated from other potential foreground objects by excluding those foreground  
15 objects that are not moving at a pace that is typical of human locomotion (e.g., 80 to 145  
16 steps per minute).

17       The output of the preprocessing step 402 defines a series of binary silhouettes,  
18 such as the collection 602 of silhouettes shown in Fig. 6. Bounding boxes can be used to  
19 enclose respective binary silhouettes. For instance, note exemplary binary silhouette 604  
20 that is enclosed by bounding box 606. The width ( $w$ ) 608 of the bounding box 606 is  
21 determined by the maximum span of the binary silhouette 604 in an x direction of the  
22 video frame. The height ( $h_b$ ) 610 of the bounding box 606 is determined by the  
23 maximum span of the binary silhouette 604 in a y direction of the video frame. Note that  
24 the respective widths of the bounding boxes in the sequence 602 change slightly as the  
25 subject 102 walks along the path 104. This is due to the fact that the respective

1 maximum spans of the binary silhouettes change over time in the x dimension as a  
2 function of the subject's 102 movement.

3 The second general step 404 of the method 400 shown in Fig. 4 uses the  
4 collection 602 of binary silhouettes shown in Fig. 6 to compute the gait-related features  
5 of stride and cadence. More specifically, cadence ( $C$ ) (in steps per minute) and stride  
6 length ( $L$ ) (in meters) are defined as:

$$7 \quad C = \frac{120 \cdot F_s}{T} \quad (1)$$

$$8 \quad L = \frac{W}{n/T} \quad (2)$$

9  
10  
11 where  $F_s$  is the frame rate (in frames per second) of the camera 106,  $T$  represents the  
12 period of the subject's gait (in frames per cycle),  $W$  represents a distance traveled along  
13 the path 104 (in meters),  $n$  represents a number of frames used in performing the analysis,  
14 and  $n/T$  is the number of gait cycles that the subject 102 traverses over the  $n$  frames. To  
15 calculate  $C$  and  $L$ , the method 400 first determines gait period  $T$  and distance walked  $W$ .  
16 To this end, substep 414 performs the task of computing the period of the subject's 102  
17 gait  $T$ , and substep 416 performs the task of computing the distance walked by the subject  
18 102.  
19

20 To begin, the substep 414 computes gait period  $T$  based on an analysis of  
21 periodicity in the bounding box width  $w$  as a function of time (e.g.,  $w(t)$ ). More  
22 specifically, human gait defines a repetitive phenomenon, so that the appearance of the  
23 walking subject 102 in a series of video frames also should exhibit periodic behavior.  
24 There are different ways to measure this periodic behavior. One technique is to calculate  
25 the width of the bounding boxes (as discussed in connection with Fig. 6) and then plot the

1 width of the bounding boxes as a function of time. Upper graph 702 shown in Fig. 7  
2 shows such a plot of bounding box width ( $w$ ) verse time ( $t$ ). The periodic appearance of  
3 the waveform shown in the upper graph 702 reflects the periodic behavior exhibited by  
4 gait.

5 More specifically, to compute the gait period  $T$ , substep 414 employs an  
6 autocorrelation procedure. In this procedure, step 414 first smoothes the width series  
7  $w(t)$ , and then piecewise detrends it to account for depth changes. Then, substep 414  
8 computes its autocorrelation,  $A(l)$ , for  $l \in [-lag, lag]$ , where  $lag$  is chosen such that it is  
9 much larger than the expected period of  $w(t)$ . The lower graph 704 of Fig. 7 shows the  
10 autocorrelation function  $A(l)$ . The peaks of  $A(l)$  correspond to integer multiples of the  
11 period of  $w(t)$ . Thus, substep 414 estimates  $\tau$ , the period of  $w(t)$ , as the average distance  
12 between every two consecutive peaks.

13 The gait period  $T$  is either equal to the width period (that is,  $T = \tau$ ) for non-  
14 fronto-parallel sequences, or equal to twice the width period (e.g.,  $T = 2\tau$ ) for fronto-  
15 parallel sequences (or near fronto-parallel sequences). More specifically, when the  
16 subject 102 walks parallel to the camera 106's imaging plane, gait appears bilaterally  
17 symmetrical (that is, the left and right legs are almost indistinguishable in this position),  
18 which result in appearance of two peaks in  $w(t)$  in each gait period, corresponding to  
19 when either one leg is leading and is maximally apart from the other. However, as the  
20 camera viewpoint departs from the fronto-parallel orientation, one of these two peaks  
21 decreases in amplitude with respect to the other, and eventually becomes  
22 indistinguishable from noise. The above-described phenomenon is evident upon a  
23 comparison of the waveforms shown in Fig. 7 and Fig. 8. More specifically, Fig. 7  
24 shows a width series (702) and its associated autocorrelation function (704) for a fronto-  
25 parallel sequence, while Fig. 8 shows a width series (802) and its autocorrelation (804)

1 for a non-fronto parallel sequence. The gait period  $T$  is defined as  $T = 2\tau$  for the case in  
2 Fig. 7 (for a fronto-parallel sequence), and defined as  $T = \tau$  for the case of Fig. 8 (for a  
3 non-fronto-parallel sequence). Note that, in accordance with the above discussion, the  
4 waveforms in Fig. 7 include more prominent peaks than the waveforms in Fig. 8.

5 One way of disambiguating the above-described two cases (for fronto-parallel  
6 and non-fronto-parallel sequences) is based on the fact that natural cadences of human  
7 walking lie in the range of 90 to 130 steps/min. Therefore,  $T$  must lie in the range  $0.923$   
8  $F_s$  and  $1.333 F_s$  frames/cycle. Since  $\tau$  and  $2\tau$  cannot both be in this interval, substep 414  
9 selects the value that is within this interval. Another way to distinguish frontal and non-  
10 frontal sequences is by analyzing the direction of motion of the person with respect to the  
11 ground plane and the calibrated camera direction.

12 As mentioned above, substep 416 shown in Fig. 4 computes the distance traversed  
13 by the subject 102, that is, the "distance walked"  $W$ . To begin with, substep 416 requires  
14 a technique for mapping between video image space and real world space. To this end,  
15 assume that the subject 102 walks on a plane and that the camera 106 is calibrated with  
16 respect to this plane (as represented in Fig. 4 by step 418). In this case, substep 416  
17 computes the 3D position  $(X_F, Y_F, Z_F)$  as the "inverse" projection of the location of the  
18 subject's feet in the image. Since the subject's 102 feet are mostly apart during walking,  
19 substep 416 uses a point half-way between the two feet as the subject's ground position  
20 in the image, denoted as  $(x_F, y_F)$ . For instance, location 612 shown in Fig. 6 is  
21 representative of an exemplary mid-feet point. Given camera intrinsic ( $K$ ) and extrinsic  
22 ( $E$ ) matrices, and the parametric equation of the plane of motion,  
23  $P : aX + bY + cZ + d = 0$  in a well-defined world frame, and assuming perspective  
24 projection:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = K \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} E \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3)$$

then we have

$$\begin{pmatrix} k_{11} & 0 & -x_F + k_{13} \\ 0 & k_{22} & -y_F + k_{23} \\ \hat{a} & \hat{b} & \hat{c} \end{pmatrix} E \begin{pmatrix} X_F \\ Y_F \\ Z_F \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -\hat{d} \end{pmatrix} \quad (4)$$

which is a linear system of three equations and three unknowns, where  $(\hat{a}, \hat{b}, \hat{c}, \hat{d}) = (a, b, c, d) \cdot E^{-1}$  and  $k_{ij}$  is the  $(i, j)$ th element of  $K$ . This system lacks a unique solution only when the subject 102 is walking directly towards or away from the camera 106 (e.g., along the optical axis of the camera), which is not the case in typical surveillance applications with cameras mounted well above the ground plane. The camera calibration parameters (e.g., the matrices  $K$  and  $E$ ) can be estimated by first obtaining pairs of corresponding image points and 3D points in the scene, then applying an exemplary calibration method described, for instance, by R. Tsai, "An Efficient and Accurate Camera Calibration Technique for 3d Machine Vision," in Proceedings of the Computer Vision and Pattern Recognition, 1986, or using some other calibration technique.

Given the above mapping strategy in equation (4), substep 416 can determine the distance walked  $W$  as the total distance traveled between a first and last 3D position on the ground plane, that is  $W = \|P_n - P_1\|$ , where  $P_1$  and  $P_n$  define the first and last positions in the subject's 102 course, respectively.

1 Having computed the gait period ( $T$ ) and the distance walked ( $W$ ), the second  
2 general step 404 now proceeds to calculate stride length ( $L$ ) and cadence ( $C$ ). More  
3 specifically, substep 422 computes stride using equation (2) defined above, while substep  
4 420 computes cadence using equation (1) defined above.

5 According to equations (1) and (2), the relative uncertainties in  $L$  and  $C$  satisfy:  
6  $\frac{\sigma_L}{L} \cong \sqrt{(\sigma_W/W)^2 + (\sigma_T/T)^2}$  and  $\frac{\sigma_C}{C} \cong \frac{\sigma_T}{T}$ , where  $\sigma_\zeta$  generally denotes the absolute  
7 uncertainty in any estimated quantity  $\zeta$ . Thus, to minimize errors, it is necessary to  
8 minimize  $\frac{\sigma_W}{W}$  and  $\frac{\sigma_T}{T}$ . This goal can be achieved by estimating  $C$  and  $L$  over a  
9 sufficiently long sequence. More specifically,  $\frac{\sigma_T}{T}$  can be reduced by making  $n$  (the  
10 number of frames in an analysis) sufficiently large. It is possible to compensate for a  
11 large  $\sigma_W$  by making  $W$  sufficiently large.

12 Having computed stride length  $L$  and cadence  $C$ , the method 400 shown in Fig. 4  
13 is now in a position to classify the subject 102 based on  $L$  and  $C$ . The third general step  
14 406 in Fig. 4 performs this task. By way of overview, an exemplary objective of the third  
15 general step 406 is to build a supervised pattern classifier that uses the cadence and stride  
16 length as input features to identify or verify a person in a given database (of training  
17 samples). Substep 426 represents the training of the model used to perform the  
18 classification, while substep 428 represents the classification of the subject 102 using the  
19 trained model developed over a course of time. That is, substep 428 determines or  
20 verifies the subject's 102 identity based on parametric classification of a cadence and  
21 stride feature vector received from the second general processing step 404.

22 More specifically, the third general step 406 can employ a Bayesian decision  
23 approach to classifying subjects based on cadence and stride length. Two different  
24 parametric models are used to model the class conditional densities in the Bayesian  
25 approach. In the first model, the cadence and stride length of any one subject are related

1 by a linear regression. In the second model, the cadence and stride length are assumed to  
2 vary as a bivariate Gaussian.

3 The task of estimating parameters for the above-described models is described as  
4 follows. Generally, a training sample of a subject's stride length and cadence is used,  
5 that is,  $(C_1, L_1), (C_2, L_2), \dots, (C_n, L_n)$ . Maximum Likelihood (ML) estimation is used to  
6 compute the model parameters of the corresponding class conditional densities. In the  
7 case of the linear regression model, stride length and cadence are known to vary  
8 approximately linearly for any one subject over his or her range of natural (or  
9 spontaneous) walking speeds, typically in the range 90-125 steps/minute. For instance,  
10 note Fig. 9, which shows measured stride length vs. cadence for a group of subjects.  
11 Note that the points associated with any one person (drawn with the same symbol) lie  
12 approximately in a straight line (note that the best fitting line is only shown for six of the  
13 subjects in the group). Hence, for each class (i.e., subject)  $\psi_i$  in the training set, the  
14 following linear regression model is assumed:  $L = a_i C + b_i + \varepsilon_i$ , where  $\varepsilon_i$  represents  
15 random noise. The class conditional probability of a measurement  $x \equiv (LC)$  is then given  
16 by:  $\Pr(x | \psi) = p_{\varepsilon_i}(r)$ , where  $p_{\varepsilon_i}$  is the probability density of  $\varepsilon_i$  and  $r = L - a_i C - b_i$  is  
17 the residual. Assuming  $\varepsilon_i$  is white noise (that is,  $\varepsilon_i \sim N(0, \sigma_i)$ ), the ML-estimate of the  
18 model parameters  $a_i$  and  $b_i$  and are obtained via a linear least squares (LSE) technique on  
19 the given training sample. Furthermore, the log-likelihood of any new measurement  $x$   
20 with respect to each class  $\psi_i$  is obtained by:  
21 
$$l_i(x) = \log p_{\varepsilon_i}(r) = \frac{1}{2} \left( \frac{r}{s_i} \right)^2 + \log s_i + \frac{1}{2} \log 2\pi$$
, where  $s_i$  is the sample standard deviation  
22 of  $\varepsilon_i$ . Since the above model only holds over a limited range of cadences  $[Cmin_i,$   
23  $Cmax_i]$ , that is,  $L = a_i C + b_i$  is not an infinite line, we set  $l_i(x) = 0$  whenever  $C$  is outside  
24  $[Cmin_i - \delta, Cmax_i + \delta]$ , where  $\delta$  is a small tolerance (for instance,  $\delta = 2$  steps/min). Since  
25

1 this range varies for each subject, it is necessary to estimate it from representative  
2 training data.

3 A less complex model of the relationship between cadence and stride length is a  
4 bivariate Gaussian distribution, e.g.,  $\Pr(x|\psi) \approx N(\mu_i, \Sigma_i)$  for the  $i$ th class. The  
5 parameters of the model,  $\mu_i$  and  $\Sigma_i$ , for the  $i$ th class are estimated respectively as the  
6 sample mean  $m_i$  and sample covariance  $S_i$  of the given training sample. The log-  
7 likelihood of a new observation  $x \equiv (CL)$  with respect to the  $i$ th class is then computed as  
8 
$$l_i(x) = \frac{1}{2}(x - m_i)^T \Sigma_i^{-1}(x - m_i) + \frac{1}{2} \ln |\Sigma_i| + \frac{1}{2} 2\pi.$$

9 In summary, according to one exemplary implementation, a method for  
10 identifying an ambulatory subject based on video analysis of the subject's gait is  
11 described. The method includes: (a) acquiring a sequence of images that collectively  
12 captures the gait of the subject; (b) determining the following features from the sequence  
13 of images: (i) cadence of gait; and (ii) stride length of gait, where the cadence and the  
14 stride length are related to each other by a linear function; and (c) identifying the subject  
15 based on the subject's cadence, stride length, and based on the linear function used to  
16 relate cadence with stride length.

17 A related apparatus and computer-readable medium for implementing the above  
18 method are also described.

### 19 C. Technique for Recognizing Subjects Using Stride, Cadence, and Height

20  
21 Fig. 10 shows another exemplary method 1000 for identifying a subject based on  
22 the gait-related features of stride and cadence, as in the case of Fig. 4, and, in addition,  
23 the parameter of height. Part of the operation of the method 1000 is similar to the method  
24 400 shown in Fig. 4, and hence, emphasis will be placed in the following discussion on  
25 functionality which is unique to method 1000.



1        Like method 400, method 1000 includes three general steps corresponding to the  
2 logic modules provided in the gait analysis logic 114 of Fig. 1. Namely, in a first general  
3 step 1002, video data 108 is preprocessed to transform it into a form suitable for feature  
4 extraction. In the second general step 1004, the transformed video data 108 is processed  
5 to extract stride, cadence, and height parameters from the transformed data. In a third  
6 general step 1006, the extracted features are used to classify the subject 102, or otherwise  
7 derive some attribute pertaining to the subject 102. Each of the above-described general  
8 steps will be described in further detail below.

9        The first general step, i.e., step 1002, is similar to general step 402 of Fig. 4. It  
10 includes a first substep 1008 that entails modeling background information in the video  
11 data 108, a second substep 1010 that entails segmenting moving objects in the video data  
12 108, and a third substep 1012 that entails tracking the subject 102 within the video data  
13 108. As previously described, substeps 1008 and 1010 generally involve identifying  
14 background information in the scene represented by the video data 108, identifying  
15 foreground information in the scene, and extracting the foreground information from the  
16 background information. Once the subject 102 has been detected in a video frame,  
17 substep 1012 involves tracking the subject 102 in subsequent frames. In one exemplary  
18 technique, binary silhouettes are tracked from frame to frame in the video data 108 using  
19 a spatial and temporal coherence technique.

20        The output of the first general step 1002, as in the case of method 400, consists of  
21 a sequence of binary silhouettes. The second general step 1004 computes cadence and  
22 stride length based on these binary silhouettes in the same manner as described above in  
23 connection with Fig. 4. Namely, substep 1014 computes gait period  $T$ , and substep 1016  
24 computes walked distance  $W$ . Substep 1018 computes cadence using equation (1)  
25 defined above, and substep 1020 computes stride length using equation (2) defined

1 above. Further, in order to properly perform the above-described steps, step 1022 is  
2 performed to calibrate the camera 106 with respect to the ground plane. More  
3 specifically, this calibration operation facilitates the estimation of  $W$  in substep 1016, as  
4 well as other operations to be described below.

5 The height parameter is measured in substeps 1024 and 1026. By way of  
6 introduction, human walking involves rhythmic up-and-down displacement of the upper  
7 body (from pelvis to head), which causes apparent bobbing of the head. Thus, the  
8 “apparent height” of the walking subject person can be modeled as a sinusoidal curve:

$$h(t) = \mu_h + \alpha_h \sin(\omega t + \phi) \quad (5).$$

11  
12 The maximum apparent height,  $\mu_h + \alpha_h$ , occurs at the mid-stance phase of walking (when  
13 the legs are closest together), and is slightly smaller than the person's height while  
14 standing still, referred to as “stature”  $H$ . The minimum height,  $\mu_h - \alpha_h$ , occurs at the  
15 mid-swing phase of walking (when the legs are furthest apart).

16 Substep 1024 computes the subject's apparent height  $h$  over a period of time  $t$ . It  
17 performs this task by mapping the height of the subject in the image,  $h_I$ , to their actual  
18 real-world apparent height  $h$ . More specifically, three different equations apply for  
19 mapping  $h_I$  to actual apparent height  $h$ , that is, for a first case where the subject 102 is  
20 entirely above the image center, a second case where the subject 102 is entirely below the  
21 image center, and a third case where the subject 102 is located at the center of the image.  
22 Fig. 11 graphically illustrates the orientations of the camera 106 and subject 102 with  
23 respect to the first, second, and third cases. Figs. 12-14 illustrate geometrical  
24 considerations relevant to the mapping of  $h_I$  to apparent height  $h$  for the above-described  
25 three cases, on which the following three equations are based:

$$h = \frac{\frac{h_l}{f} Z}{\cos \theta_v - \frac{y_u}{f} \sin \theta_v} \quad (6)$$

$$h = \frac{\frac{h_l}{f} Z}{\cos \theta_v - \frac{y_b}{f} \sin \theta_v} \quad (7)$$

$$h = \frac{\frac{h_l}{f} Z}{\cos \theta_v - \frac{y_u}{f} \sin \theta_v} + h = \frac{\frac{h_l}{f} Z}{\cos \theta_v - \frac{y_b}{f} \sin \theta_v} \quad (8)$$

where  $y_b$  and  $y_u$  are coordinates in the image respectively defining the subject's lower body (e.g., feet) and upper body (e.g., head),  $h_l = y_u - y_b$  is the subject's height in the image,  $\theta_v$  is the tilt angle of the camera 106,  $f$  is the camera focal length (in pixels), and  $Z$  is the distance from the camera center to the subject (that is, the depth). Again, the above three equations pertain to the cases where the subject is entirely above the image center (equation 6), entirely below the image center (equation 7), or neither completely above or below the image center (equation 8). Note that when the person is sufficiently far from the camera,  $h$  reduces to the numerator  $\frac{h_l}{f} Z$ . As indicated by step 1022 in Fig. 10, proper calibration of the camera with respect to the ground plane should be performed to ensure calculation of apparent height.

Figs. 15-18 further clarify the meaning of the above-identified parameters used to calculate apparent height. With reference first to the stick figure model of Fig. 15, the subject's height in the image,  $h_l$ , is estimated as the vertical distance between the head and the feet of the subject. In a non-fronto-parallel sequence, which pertains to the representation shown in Fig. 15, this distance is approximated as a vertical segment 1502

that extends from the top of the bounding box 1504 (e.g., where the head is located) to the point 1506 halfway between the subject's two feet (1508, 1510, respectively). Fig. 16 shows the application of the height measurement strategy shown in Fig. 15 to an actual sequence of binary silhouettes in a non-fronto-parallel sequence. Fig. 17 shows an exploded view of one of the binary silhouettes shown in Fig. 16, particularly showing the measurement of the x coordinates corresponding to respective positions of the subject's feet. This information can be used to find the point 1702 positioned midway between the subject's feet. Finally, Fig. 18 shows the measurement of height  $h_l$  for a fronto-parallel sequence. In the fronto-parallel case,  $h_l$  corresponds to the bounding box height ( $h_b$ ) of the binary silhouette.

Having computed the time-series  $h(t)$  of apparent heights of a walking person measured over a video sequence of length  $n$ , substep 1026 comes into play by fitting the measured height values to the model defined above in equation (5), and then identifying the parameters which provide the best fit. More specifically, assuming a known frequency of gait  $\omega$  (note  $\omega \equiv \frac{2\pi}{T}$ ), substep 1026 estimates three model parameters using a least squares fitting technique. More specifically, assuming a data model:

$$h(t) = \mu_h + \alpha_h \sin(\omega t + \phi) + \varepsilon_t, \quad t = 0, 1, \dots, n \quad (9)$$

the unknown parameters of the model that minimize the sum-square of the residuals  $\varepsilon$  are given by  $\mu_h = \frac{1}{n} \sum_{t=1}^n h(t)$  and  $\alpha_h = \sqrt{A^2 + B^2}$ , where  $A = \frac{2}{n} \sum_{t=1}^n (h(t) - \bar{h}) \cos \omega t$  and  $B = \frac{2}{n} \sum_{t=1}^n (h(t) - \bar{h}) \sin \omega t$ . Figure 19 shows an example of a height series (dashed line 1902) fitted to a model (solid line 1904) via the above method. In this exemplary case,  $\mu_h \cong 175.3\text{cm}$  and  $\alpha_h \cong 2.7\text{cm}$ , and the subject's actual height is 177cm.

The stature ( $H$ ) of the subject 102 can be approximated relatively accurately from the apparent height parameters, that is,  $H \cong \mu_h + \alpha_h$ . However, in subsequent classification analysis, method 1000 shown in Fig. 10 uses the raw parameter data ( $\mu_h$  and  $\alpha_h$ ) rather than the stature height. This is because  $\mu_h$  and  $\alpha_h$  are together more “informative” than  $H$  for the purpose of characterizing gait behavior.

As to error analysis, as mentioned above, the uncertainty both in  $C$  and  $L$  is a decreasing function of the number of steps walked,  $N$ , and hence can be effectively reduced by using a large  $N$ , regardless of whether the uncertainty is caused by random or systematic error. Systematic error can, for example, be caused by strong shadows that are segmented with the silhouette. However, the uncertainty in height does not possess the above-described property; that is, the uncertainty in height does not necessarily decrease when estimated over many frames. Intuitively, this is because height is estimated independently in each frame, while a single estimate of cadence and stride length is computed over an entire sequence. Assuming for simplicity that  $\theta_v \approx 0$  in equation Nos. 6-8, then the uncertainty  $\sigma_{\hat{h}}$  in each height measurement  $\hat{h}$  is given by:

$$\sigma_{\hat{h}} = \hat{h} \sqrt{\left(\frac{\sigma_{h_l}}{h_l}\right)^2 + \left(\frac{\sigma_z}{Z}\right)^2} \quad (10)$$

Thus, the only way to reduce this uncertainty is by reducing the uncertainty in  $Z$  and  $h_l$ , denoted by  $\sigma_z$  and  $\sigma_{h_l}$ , respectively.

The above-described four gait features,  $\mu_h$ ,  $\alpha_h$ ,  $C$ , and  $L$ , are all correlated, and it can be expected that they together form “tight modes” in 4-D space for use in classification. To this end, general step 1006 recognizes subjects captured in video using the gait-related parameters described above, namely  $C$ ,  $L$ ,  $\mu_h$  and  $\alpha_h$  (note that apparent height can be regarded as a gait-related parameter because it is an artifact of the walking

1 movement). More specifically, the third general step 1006 builds a supervised pattern  
2 classifier that uses the two height-related parameters ( $\mu_h, \alpha_h$ ), C, and L as input features  
3 to identify or verify a person in a given database (of training samples). That is, substep  
4 1028 represents the training of the model used to perform the classification, while substep  
5 1030 represents the classification of the subject 102 using the trained model developed  
6 over a course of time. More specifically, substep 1030 determines or verifies the  
7 subject's 102 identity based on the gait-related parameters C, L,  $\mu_h$  and  $\alpha_h$  discussed  
8 above. In one implementation, the classification can use a K nearest neighbor non-  
9 parametric pattern classifier to perform its classification tasks in the 4-D feature space of  
10 the above-described four gait-related parameters.

11 In summary, according to one exemplary implementation, a method is described  
12 for determining height parameters that describe a dynamically varying height of an  
13 ambulatory subject based on video analysis of the subject. The method includes: (a)  
14 acquiring a sequence of images that collectively captures the gait of the subject; (b)  
15 measuring a dynamically varying height function of the subject based on an analysis of  
16 the varying height of the subject in the sequence of images; and (c) fitting the  
17 dynamically varying height function of the subject to a model that describes varying  
18 height, wherein the height parameters correspond to parameters used in the model. In  
19 one exemplary implementation, the height parameters, along with cadence of gait and  
20 stride length of gait, can be used to identify the ambulatory subject.

21 The above-mentioned model represents an ideal variance in height as a sinusoidal  
22 function. The parameters used in the model include a first parameter that describes a  
23 baseline height value exhibited by a person in ambulatory motion, and a second value  
24 that describes a maximum deviation from the baseline height value. The above-  
25 mentioned measuring of the dynamically varying height function includes: (a) extracting

1 a series of depictions of the ambulatory subject from a larger body of image information  
2 contained within the sequence of images; (b) defining a series of bounding boxes that  
3 enclose respective depictions; and (c) for each of the depictions, determining a distance  
4 between a point midway between the feet of the subject and a top of the depiction's  
5 associated bounding box.

6 A related apparatus and computer-readable medium for implementing the above  
7 method are also described.

#### 8 9 D. A Non-Parametric Technique for Recognizing Subjects Using Self-Similarity 10 Plots

11 Fig. 20 shows another exemplary method 2000 for identifying a subject based on  
12 gait-related features. This method 2000 relies on a holistic technique to derive the  
13 features used to characterize gait, rather than a theoretical model-based technique as in  
14 the case of the prior two methods (in Figs. 4 and 10). In particular, the method 2000 uses  
15 self-similarity plots (SSP's) (to be defined below) to extract features from video images  
16 without having to identify and extract specific model-based information from the video  
17 images. The method 2000 is deemed "holistic" because the SSP represents a direct  
18 transformation of the spatiotemporal volume (XYT) of the walking person. Otherwise,  
19 part of the operation of the method 2000 is similar to the methods 400 and 1000 shown in  
20 Figs. 4 and 10, respectively, and hence, emphasis will be placed in the following  
21 discussion on functionality which is unique to method 2000.

22 Like methods 400 and 1000, method 2000 includes three general steps  
23 corresponding to the logic modules provided in the gait analysis logic 114 of Fig. 1.  
24 Namely, in a first general step 2002, video data 108 is preprocessed to transform it into a  
25 form suitable for feature extraction. In the second general step 2004, the transformed

1 video data 108 is processed to extract gait-related parameters using self-similarity plots.  
2 In a third general step 2006, the extracted features are used to classify the subject 102, or  
3 otherwise derive some attribute pertaining to the subject 102. Each of the above-  
4 described general steps will be described in further detail below.

5 The first general step, i.e., step 2002, is similar to general steps 402 and 1002 of  
6 Figs. 4 and 10, respectively. It includes a first substep 2008 that involves modeling  
7 background information in the video data 108, a second substep 2010 that entails  
8 segmenting moving objects in the video data 108, and a third substep 2012 that entails  
9 tracking the subject 102 within the video data 108. As previously described, substeps  
10 2008 and 2010 generally involve identifying background information in the scene  
11 represented by the video data 108, identifying foreground information in the scene, and  
12 extracting the foreground information from the background information. Once the  
13 subject 102 has been detected in a video frame, substep 2012 involves tracking the  
14 subject 102 in subsequent frames. In one exemplary technique, binary silhouettes are  
15 tracked from frame to frame in the video data 108 using a spatial and temporal coherence  
16 technique.

17 The output of the first general step 2002, as in the case of method 2000, consists  
18 of a sequence of binary silhouettes. In the context of the method 2000, the binary  
19 silhouettes are referred to as “templates.” (For instance, image 502 in Fig. 5 represents  
20 one exemplary template.) Each of these templates is characterized by a bounding box  
21 which encloses a “blob” that depicts a snapshot of the walking subject 102 at a particular  
22 point in time. The bounding boxes generally resemble the bounding boxes (e.g., box  
23 606) shown in Fig. 6. However, in this application, the bounding boxes may include an  
24 “empty” region around its perimeter (e.g., a 10-pixel border) in order to allow for shifting  
25 that may occur when later computing the cross-correlation of template pairs (to be



described below). In alternative implementations, the templates 502 or 506 shown in Fig. 5 can be used to extract gait-related parameters, that is, as opposed to the binary template 508. There are competing tradeoffs to using any of the image templates shown in Fig. 5 in measuring image similarity for computing the SSP. For instance, the first template 502 does not require segmentation and is hence insensitive to segmentation errors. The third template 508 is more robust to clothing and background variations. And the second template 506 provides a hybrid of the first and third kinds of templates (502, 508); that is, the second template 506 is robust to background variations but sensitive to segmentation errors and clothing variations.

The general step 2004 extracts features from the above-described templates. This step 2004 begins with substep 2014, which entails aligning and scaling the templates. More specifically, template size can vary according to camera viewpoint and depth. Accordingly, substep 2014 scales the templates to a fixed height (e.g., 50 pixels) prior to computing the SSP, and also aligns the templates.

Substep 2016 computes a self-similarity plot (SSP) based on the collected templates. One exemplary technique for forming the SSP is as follows. Let  $I_i$  represent the  $i$ th scaled template with size  $\tilde{w}_i \times \tilde{h}_i$  (in pixels). The corresponding self-similarity plot  $S(i, j)$  is computed as the absolute correlation of each pair of templates  $I_i$  and  $I_j$ , minimized over a small search radius  $r$ , namely:

$$S(i, j) = \min_{|dx| < r, |dy| < r} \sum_{|x| \leq \frac{W}{2}} \sum_{|y| \leq \frac{H}{2}} |I_j(x + dx, y + dy) - I_i(x, y)| \quad (11)$$

where  $W = \min(\tilde{w}_i, \tilde{w}_j - 2r)$  and  $H = \min(\tilde{h}_i, \tilde{h}_j - 2r)$  so that the summation does not go out of bounds. In other words, the SSP constitutes a matrix of cross-correlation between pairs of image templates in the sequence of templates. The term “self-similarity” denotes

the fact that the sequence of templates is compared with itself to provide a measure of similarity between different parts of the sequence of templates. Absolute correlation is used in the above analysis for its simplicity. But other similarity measures can be used, including normalized cross-correlation, the ratio of overlapping foreground pixels, Hausdorff distance, Chamfer distance, binary thresholding of the image (which removes variations of clothing colors and only uses shape and motion to compute the SSP), etc.

Figs. 21 and 22 highlight some of the properties of  $S$  for fronto-parallel (Fig. 21) and non-fronto-parallel (Fig. 22) walking sequences. The horizontal axis in an SSP represents an arrangement of templates in the order in which these templates were captured by the video camera (e.g., as in the sequence 602 shown in Fig. 6). The vertical axis in an SSP represents a sequence of the very same templates. Accordingly, any point in the SSP pairs one template in the sequence with another template in the same sequence (the diagonal running down the middle of the SSP identifying two versions of the same template in the sequence). In Figs. 21 and 22, similarity values computed using equation (11) are linearly scaled to the grayscale intensity range (0 to 255) for visualization.

More specifically, the dark diagonals in Figs. 21 and 22 are due to the periodicity of gait (e.g., note exemplary diagonal 2102 in Fig. 21 and exemplary diagonal 2204 in Fig. 22), while the cross-diagonals are due to the temporal mirror symmetry of the gait cycle (e.g., note exemplary cross-diagonal 2104 in Fig. 21 and exemplary cross-diagonal 2202 shown in Fig. 22). The intersections of these diagonals, e.g., representing the local minima of  $S$ , correspond to key poses of the gait cycle. Specifically, each intersection corresponds to a combination of the following four key poses of gait: (i) when the two legs of the subject are furthest apart and the left leg is leading; (ii) when the two legs are joined together and the right leg is leading; (iii) when the two legs are furthest apart and the left leg is leading; and (iv) when the two legs are joined together and the left leg is

1 leading. These poses are labeled as A, B, C, and D, respectively. Figs. 21 and 22 include  
2 exemplary templates that illustrate the poses corresponding to labels A, B, C, and D.  
3 That is, note exemplary poses 2106 in Fig. 21, and exemplary poses 2206 in Fig. 22.  
4 Some of the intersections shown in Fig. 21 are not present in the SSP of Fig. 22 because  
5 the non-fronto-parallel gait does not appear bilaterally symmetric, and thus does not give  
6 rise to these diagonals. Generally, it can be seen that the SSP encodes both the frequency  
7 and phase of the gait cycle.

8 Fig. 23 shows the formation of a collection of self-similarity units (SSU's) on an  
9 SSP. Each SSU, such as exemplary SSU 2306, defines a sub-matrix of the SSP that starts  
10 at a defined phase in the sequence of templates and spans one gait cycle. The SSP can  
11 thus be viewed as a tiling of (contiguous) SSU's. A different tiling can be obtained by  
12 selecting a different starting phase. Fig. 23 shows the use of SSU's corresponding to the  
13 left and right double-support poses for gait recognition (e.g., defined by lines 2302 and  
14 2304). Because the SSP is approximately symmetric, and to expedite analysis, the  
15 method 2000 only uses the SSU's from the top half of the SSP (or the bottom half), as  
16 shown in Fig. 23 (that is, the upper triangular matrix of the SSP or the lower triangular  
17 matrix). For a sequence containing K gain cycles, there are  $2 \frac{k(K+1)}{2} = K(K+1)$   
18 potential SSU's. The SSU's constitute discrete chunks of information that encapsulate  
19 information regarding the subject's gait, and are the counterpart to the discrete model-  
20 based features extracted in methods 400 and 1000.

21 Substep 2018 in Fig. 20 entails the extraction of features from the self-similarity  
22 information and the normalization of such features. The extraction of features can  
23 involve extracting gait frequency and phase via analysis of the SSP, which, as mentioned  
24 above, inherently encodes this information in its patterns. That is, substep 2108 applies  
25 an autocorrelation method on the SSP. This technique first smoothes the autocorrelation

1 matrix of the SSP, computes its peaks, and then finds the best-fitting regular 2D lattice  
2 for these peaks. The period is then obtained as the width of this best-fitting lattice. The  
3 phase is computed by locating the local minima of the SSP that correspond to the A and  
4 C poses (depicted in Figs. 21 and 22). The above-described extracted information can  
5 assist in defining the boundaries of the SSU's described above.

6 The objective of normalization is to equalize different attributes of the SSU's to  
7 facilitate effective comparison of such information in the classification step 2006. More  
8 specifically, in any pattern classifier, it is useful to determine which sources of variation  
9 in the input data are irrelevant to classification, and extract the influence of such sources  
10 via normalization prior to classification. In the present case, an SSU of the same walking  
11 subject will vary with at least: (i) clothing; (ii) the background scene; (iii) number of  
12 pixels on target; (iv) camera viewpoint; and (v) walking speed. It is useful to attempt to  
13 neutralize one or more of these factors prior to classification so that the classification  
14 results are not unduly influenced by these factors. Normalization can also include scaling  
15 the SSU's to a fixed size, etc.

16 The SSU output of the feature extraction general step 2004 contains a wealth of  
17 information regarding the characteristics of the walking subject. However, this  
18 information set may be too large to perform effective classification on. Hence, the  
19 classification step 2006 can include a substep 2020 of reducing the amount of  
20 information presented to it from the feature extraction step 2004. Different techniques  
21 are available for performing this data reduction. In one technique, principal component  
22 analysis is performed to analyze the features identified by the feature extraction step 2004  
23 to identify a reduced set of information that best characterizes the behavior of the walking  
24 subject. The principal component analysis effectively translates a collection of features  
25 from a first "space" to a smaller second "space." In other words, the second space has a

1 smaller "dimension" than the first space. Because the features in the second space have  
2 been selected for their descriptive merit, these features are analogous to the mathematical  
3 concept of eigenvalues; accordingly, the space that contains such salient features can be  
4 informally referred to as a kind of "eigengait" space. Principal component analysis is  
5 well known in the art, and is described in an image processing context in, for example,  
6 M. Turk and A. Pentland, "Face Recognition Using Eigenfaces," in CVPR, 1991.

7 Finally, substep 2022 determines the identify of a subject under consideration by  
8 comparing the reduced feature set with prior stored feature sets pertaining to a plurality of  
9 different subjects. Any kind of pattern recognition technique can be used to perform this  
10 task, such as, but not limited to, K nearest neighbor analysis. In the context used here,  
11 each SSU can be represented as a d-dimensional vector by concatenating its m rows. The  
12 distance between two SSU's can then simply be computed as their Euclidean distance in  
13 this space to determine whether vector under consideration is a good match for  
14 previously stored vectors.

15 A method for identifying an ambulatory subject based on video analysis of the  
16 subject's gait is described. The method includes: (a) acquiring a sequence of images that  
17 collectively captures the gait of the subject; (b) extracting a series of depictions of the  
18 ambulatory subject from a larger body of image information contained within the  
19 sequence of images; (c) determining a self-similarity plot based on the extracted series of  
20 depictions; (d) determining a series of self-similarity units within the self-similarity plot,  
21 the self-similarity units associated with one gait cycle of the subject's ambulatory motion;  
22 and identifying the subject based on the self-similarity units.

23 The above-described method further includes determining relevant features  
24 associated with the self-similarity units, where the identifying the subject based on the  
25 self-similarity units includes identifying the subject based on the determined relevant

1 features. The above-mentioned determining of relevant features includes performing  
2 principal component analysis to identify the relevant features. The above-mentioned  
3 identifying includes performing K nearest neighbor analysis to determine the identity of  
4 the subject.

5 A related apparatus and computer-readable medium for implementing the above  
6 method are also described.

#### 7 8 E. Techniques for Recognizing Subjects Combining Parametric and Holistic 9 Approaches

10 The above-described classification techniques used either a parametric-based  
11 approach or a holistic based approach, but not both. This section describes exemplary  
12 ways in which parametric analysis can be combined with holistic analysis to potentially  
13 enhance the accuracy of gait recognition.

14 Fig. 24 represents one such method 2400 for combining parametric analysis with  
15 holistic analysis. This method 2400 involves a step 2402 of extracting and normalizing a  
16 set of features based on a model-based parametric approach. (Note that the preprocessing  
17 operations that precede the feature extraction operation have been omitted from Fig. 24 to  
18 simplify the discussion). An exemplary output of step 2402 can consist of parameters  
19 describing stride length ( $L$ ), cadence ( $C$ ), and height ( $\alpha_h, \mu_h$ ), etc. The method 2400 also  
20 involves a step 2404 of extracting and normalizing a set of features based on a holistic  
21 approach. An exemplary output of the step 2404 can consist of a collection of SSU's  
22 contained within an SSP.

23 Step 2406 generally represents the combination of the features provided by steps  
24 2402 and 2404. In the example presented in Fig. 24, the output of step 2406 can consist  
25 of the features of  $L, C, \alpha_h, \mu_h$ , and a collection of SSU's.

1 Step 2408 performs pattern recognition based on the set of features supplied to it  
2 by step 2406. Any kind of classification strategy can be used in step 2408, such as the K  
3 nearest neighbor approach.

4 Fig 25 shows another method 2500 for combining parametric analysis with  
5 holistic analysis. This method 2500 includes a step 2502 for using parametric analysis to  
6 identify a list of L likely matches between a subject under consideration and subject's  
7 previously registered in the knowledgebase of the pattern recognizer. Step 2504  
8 performs holistic analysis on data associated with the list of L potential candidates to  
9 select a most closely matching candidate. For example, by extracting the cadence, stride  
10 length, and height parameters, step 2502 can identify a list of 10 potential candidates that  
11 might match the identity of an ambulatory subject under consideration. Then, by using  
12 SSP analysis, step 2504 can reexamine the data associated with the 10 candidates to  
13 identify the candidate that most closely matches the subject under consideration.

14 In another implementation, the roles of the parametric analysis and holistic  
15 analysis are reversed from that shown in Fig. 25. In this case, holistic analysis can be  
16 used to identify a list of L potential candidates, and parametric-based analysis can be  
17 used to select the best candidate from this list.

18 In summary, according to one exemplary implementation, a method is described  
19 for identifying an ambulatory subject based on video analysis of the subject's gait. The  
20 method includes: (a) acquiring a sequence of images that collectively captures the gait of  
21 the subject; (b) determining at least one of the following parametric-derived features from  
22 the sequence of images: (i) cadence of gait; (ii) stride length of gait, and (iii) height of the  
23 subject; (c) determining a self-similarity plot based on the sequence of images to provide  
24 at least one holistic-derived feature; and (d) identifying the subject based on the at least  
25 one parametric-derived feature and the at least one holistic-derived feature.

1 A related apparatus and computer-readable medium for implementing the above  
2 method are also described.

3 According to another exemplary implementation, a method is described for  
4 identifying an ambulatory subject based on video analysis of the subject's gait. The  
5 method includes: (a) acquiring a sequence of images that collectively captures the gait of  
6 the subject; (b) determining at least one parametric-derived feature from the sequence of  
7 images; (c) identifying a group of individuals that have a prescribed probability of  
8 matching the identity of the subject based on the at least one parametric-derived feature;  
9 and (d) performing holistic gait analysis to determine a candidate in the group of  
10 individuals having the greatest assessed probability of matching the subject.

11 A related apparatus and computer-readable medium for implementing the above  
12 method are also described.

13 Still other variations on the above-described design strategy are possible.

14 In conclusion, although the invention has been described in language specific to  
15 structural features and/or methodological acts, it is to be understood that the invention  
16 defined in the appended claims is not necessarily limited to the specific features or acts  
17 described. Rather, the specific features and acts are disclosed as exemplary forms of  
18 implementing the claimed invention.